Introduction

The COVID-19 pandemic caused many workers to relocate as their jobs became locale independent. Anecdotal evidence suggests that many moved out of densely packed urban business centers like New York City and San Francisco, and it is not hard to imagine why someone may not want to live in a crowded city in the midst of a pandemic. Because anecdotal evidence has its limitations, is there a way we could use data to quantify the effect of the threat of COVID-19 on the housing market?

The following memo will describe the effect that 2020 county-level per-capita COVID-19 mortality rates had on 2021 county-level housing prices. It will investigate this effect in the context of county-level wage changes from 2020 to 2021. If our anecdotes align with the data, we would expect to see home prices fall in counties with comparatively higher COVID-19 mortality rates; this result would be of value to researchers in many fields.

Data

Investigating this question in this specific context and with reasonable controls requires reliable compilation of the following statistics:

- County-level change in housing prices from 2020 to 2021
- 2020 County-level per-capita COVID-19 mortality rates
- 2021 County-level average wages
- County-level change in average wage from 2020 to 2021

Some of these statistics are readily available; others require creativity to compile. For example, both of the types of wage data we need are available for download from the U.S. Bureau of Labor Statistics in the exact form in which we will analyze them. Unfortunately, that is where the easy stats dry up, and creative compiling enters the game.

The most reliable source of 2020 county-level population data would be the 2020 U.S. Census. However, many of the results of that survey are not yet publicly available. However, the county-level race-based population data that states use to draw new congressional districts is publicly available; that dataset provides us with reliable county-level populations from which to create the per-capita COVID-19 mortality rate that we will use in our analysis. To get the COVID-19 part of that statistic, we will use what is arguably one of the most reliable sources of 2020 county-level COVID-19 deaths: The New York Times' compilation of daily county-level COVID-19 death reports. By summing those daily county death reports into a given county's

total year COVID-19 death count, dividing that count by the population of the county, and multiplying the result by one hundred thousand, we obtain that county's 2020 per-capita COVID-19 mortality rate.

After finishing that great bout of creative compiling, a new data problem surfaces. Without 2021 housing data, we would not be able to track changes in housing prices from 2020 to 2021. The only apparent source of sub-state-level 2021 housing price data is the Federal Housing Finance Authority's Three-Digit ZIP Code Housing Price Index. This data is not organized at the county level, so it is impossible to compare to our county-level mortality rates in its unedited form. However, the Department of Housing and Urban Development provides a helpful ZIP-FIPS crosswalk – a data tool that allows us to convert ZIP codes into FIPS codes, which are unique to a given county. Using that crosswalk, we obtain county level data on the 2020 and 2021 Housing Price Index; we then create from those data a single statistic that measures the percent change from 2020-2021 in the Housing Price Index for a given county.

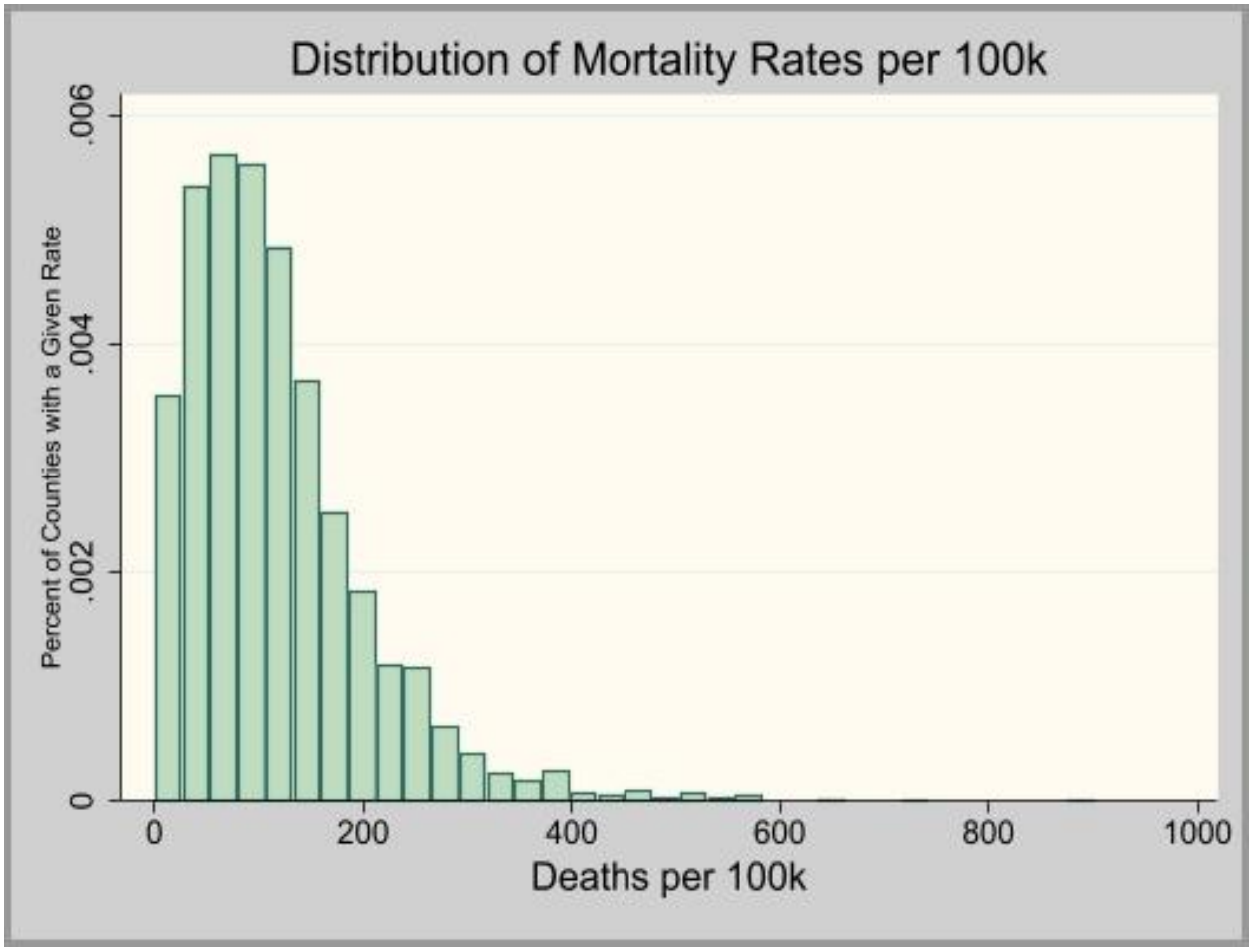
Having now compiled the statistics of interest, we can analyze them together and begin to form an idea of how they interact. We have included some notable early observations below:

Observation	**County**	**Statistic**
Largest increase in housing prices	Ada County, ID	+18.5%
Largest decrease in housing prices	San Francisco County, CA	-4.2%
Largest increase in average weekly wage	Lunenburg County, VA	+175.2%
Largest decrease in average weekly wage	Thomas County, NE	-30.1%
Highest COVID-19 mortality per 100k	Jerauld County, SD	902.0
Lowest COVID-19 mortality per 100k	93 Different Counties	0

County Level Summary Data

Observation	**Mean**	**Standard Deviation**
Average change in county housing prices	+6.959%	2.301
Average change in county weekly wage	+2.782%	6.113
Average county deaths per 100k	117.334	89.126

Population Unweighted Averages Across All Counties



Histogram of Mortality Rates Across Counties Showing Heavy Left Concentration of the Same

Empirical Analysis

The individual COVID-19 mortality rate of any given county cannot completely account for a change in that county's overall housing prices. Many other factors come into play, including inflation and regional differences in wealth, among others. In order to control for those other factors, we use an ordinary least squares regression model with several independent variables. This allows us to determine the strength of the relationship between housing prices and COVID-19 mortality rates while also controlling for things like overall inflation and regional differences. The specific regression model that we use in this analysis is as follows:

$$housechange_i = ß_0 + ß_1 log(mortper100k)_i + ß_2 log(avgwage)_i + ß_3 wagechange_i + u_i$$

…where *housechange* represents a given county's percent change in HPI from 2020 to 2021, *log(mortper100k)* represents that county's COVID-19 mortalities per one hundred thousand residents in 2020, *log(avgwage)* represents that county's average weekly wage in 2021, *wagechange* represents that county's percent change in average weekly wage from 2020 to 2021, *u* represents an error term that accounts for all variation in *housechange* not measured by those variables, and the beta coefficients $ß_i$ represent the effect of a one unit increase of a given variable on *housechange*.

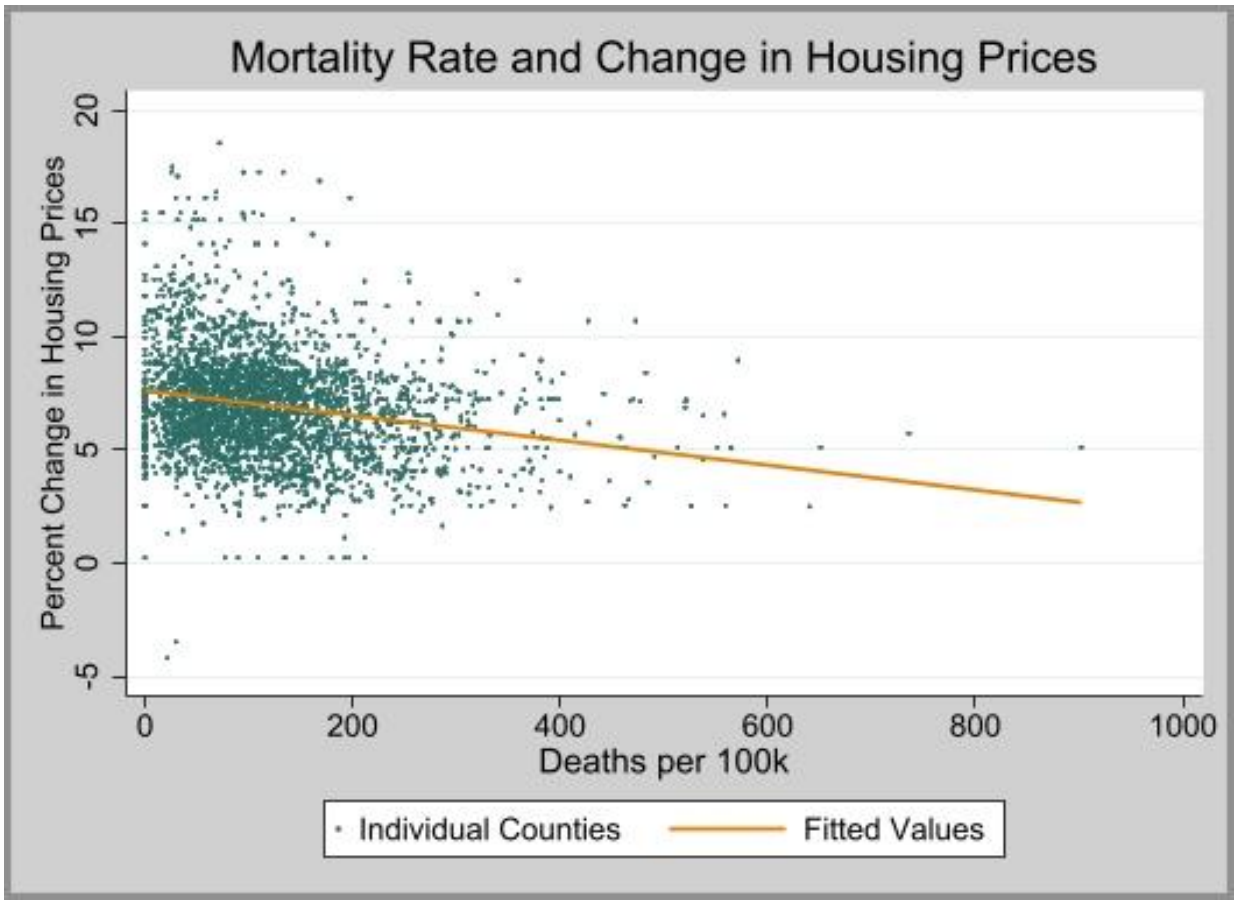
This model relies on a number of assumptions: first, it assumes that measures of county wages and changes in wages do a good job of controlling for overall inflation and regional differences in county wealth. We argue that controlling for average weekly wage allows our model to adequately capture the effects of living in a poorer or richer region of the country without needing to explicitly specify those regions. We also argue that controlling for year-on change in average weekly wage allows our model to adequately capture the effects of inflation and COVID-related labor market adjustments, sufficiently isolating the effect of COVID-19 mortality rates on changes in housing prices from other outside influences to allow for causal analysis. Running the above-specified regression with standard robustness checks returned the following:

Variable	**Coefficient**	**Std. Error**	**t-score**	**P>|t|**
log(mortper100k)	-0.6947	0.0547	-12.70	0.000
log(avgwage)	-0.6442	0.2535	-2.54	0.011
wagechange	0.0311	0.0111	2.81	0.005
_constant	14.3614	1.7547	8.18	0.000
Observations: 3035			R-squared:	0.0653

Regression Results Table

The coefficients on each variable (other than the constant) give us the elasticity of housing price change relative to each variable, which means that a one percent increase in a given variable would predict a $ß_i$ percent change in housing prices. The t-scores and P-values are statistical tools that measure the significance of each of our coefficients; in this case, they warrant no mention outside of showing that our coefficients are all statistically significant.

In addition to giving us the coefficients for each of our variables, our regression analysis also allows us to predict housing change in a given county if we know its measures of our other variables. This also allows us to plot those predictions and visualize them in the context of our individual data points, as pictured below:



Scatter Plot of Mortality Rates and Changes in Housing Prices with Regression-predicted Values

The results from our regression analysis establish what we argue is a causal relationship between a county's 2020 COVID-19 mortality rate and its change in housing prices from 2020 to 2021. Let us consider Example County, ZX and Sample County, ZX, which are equal in every statistic except for COVID-19 mortality rate. Our results suggest that if Example County had a 10% higher 2020 COVID-19 mortality rate than Sample County, then any housing price increase in Example County would be 6.95% smaller than the increase in Sample County. That means that a house that would have sold for $300,000 in both counties in 2020 could theoretically sell

for $350,000 in Sample County (a 16.7% increase from 2020 to 2021), but an identical house in Example County could only sell for about $334,750 (an 11.58% increase from 2020 to 2021).

In a year in which housing prices rose in almost every county in the country, these results show that counties with relatively high COVID-19 mortality rates experienced that nationwide increase to a significantly lesser degree than counties with relatively low COVID-19 mortality rates. Lower prices can imply lower demand, and there may have been lower demand for housing in high COVID-19 mortality rate counties because displaced workers searching for new housing were substituting toward counties with low COVID-19 mortality rates.

One potential weakness of this interpretation involves population density. At a gut-check level, the counties with lower COVID-19 mortality rates seem to belong to rural areas like Idaho and South Dakota with sparsely distributed populations. Given that higher population density would make spreading a disease like COVID-19 easier, making high-density counties more likely to have a relatively high COVID-19 mortality rate, it stands to reason that the large predicted effect that COVID-19 mortality rates have on county house prices in our model may have been erroneously inflated by the exclusion of population density as a control. For this exact reason, however, it may be the case that COVID-19 mortality rates serve as an instrument for population density in this model. If this is true, then our interpretation would still hold in a modified form: displaced workers may have substituted to rural counties (with low COVID-19 mortality rates) in pursuit of less crowded living, and that attractive population sparsity may have caused those rural counties to have lower COVID-19 mortality rates.

Conclusion

Our analysis suggests that there may be a causal link between a given county's COVID-19 mortality rate in 2020 and the change in its housing prices from 2020 to 2021. While we have suggested one possible causal interpretation and explored a potential modification of that interpretation, more thorough analysis with standard controls for regional, racial, and density differences across counties may yield different results entirely. Therefore, this particular analysis should be limited to its defined scope, and understood to be an introduction to a subtopic of COVID-19 research that deserves further, more thorough investigation.

Supporting Materials